



Data Pedigree in the Collaboratory for Multi-scale Chemical Science (CMCS)

Carmen Pancerella

carmen@ca.sandia.gov

Sandia National Laboratories

June 2002





CMCS Pedigree

OBJECTIVE Data pedigree is a relationship which provides a “line of ancestors”. This allows for the categorization and tracing of the scientific data, possibly across scales and for the identification of the data’s ultimate origin. Pedigree data and metadata will be associated with CMCS resources, and will be browsable and searchable from the CMCS portal.

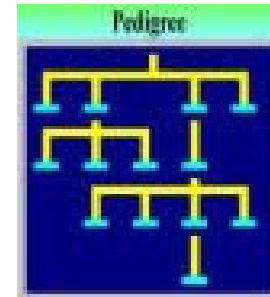
TASK This task involves both application scientists to define metadata and to determine what pedigree information is necessary and infrastructure team members to implement the solutions.

TEAM Carmen Pancerella (lead), Jim Myers, Brett Didier, Eric Stephan, John Hewson, Tom Allison, Branko Ruscic, David Leahy (honorary member), Michael Frenklach (honorary member)





Why is pedigree critical to CMCS?

- CMCS data will be identifiable and traceable.
- With a pedigree browser, a scientist can
 - › trace scientific data to its ultimate origin, possibly across scales.
 - › track data to program versions, and hence, to program bugs reported for that version.
 - › retrieve data and get papers/reports about this data.
 - › retrieve data and find references about its input data and parameters.
 - › retrieve data at one scale and trace input data to techniques at a lower scale.





Pedigree Tasks

- ☒ Document requirements for pedigree.
- ☒ Create tutorial for CMCS team.
- ☒ Survey metadata standards, in particular, Dublin Core. Make recommendations for use of Dublin Core in the CMCS.
-  Define standard mechanism to represent subject-verb-object triples in DAV properties.
-  Append data with pedigree information in WebDAV. At this time, GRIMech data (Michael Frenklach), HCT data (Bill Pitz), and EMSL Computation Results database (Theresa Windus) are on the Slide server and annotated with some pedigree metadata.



Pedigree Tasks



Develop portlets to facilitate insertion and modification of pedigree metadata.



Develop pedigree browser portlet.



To Do List

Develop pedigree API to facilitate insertion and modification of pedigree metadata.



To Do List

Develop pedigree API for pedigree browsing.



Portlet for Adding Resources and Pedigree Information Into CMCS



Collaboratory for Multi-Scale Chemical Science

Welcome
**Tommy
Turbine**
Customize:
[HTML](#) [WML](#)
[Edit](#)
[account:](#)
[turbine](#)
[Logout](#)

About CMCS

Portal

Project Team

Publications

Contacts

[Home](#) [CMCS Search](#) [Pedigree](#) [Data Insertion](#)

CMCS Put File and Pedigree Data Portlet

Putting Resources with Pedigree Information into CMCS WebDAV Repository

WebDAV Repository

WebDAV Username

WebDAV Password

Upload File

Author/creator

Other Contributors (separated by semicolons)

Resource Name

Subject and Keywords (separated by)



Pedigree Metadata Stored as WebDAV Properties



DAV Explorer

File Edit View Help

http:// cmcs.ca.s

DAV Explorer

- http://cmcs.ca.sand
- \

Name

- DublinCoreWebDAV
- PedigreeTutorial.doc**
- Zope.doc

View/Modify Properties

cmcs.ca.sandia.gov:18081/slide/users/carmen/PedigreeTutorial.doc

Tag	Namespace	Value
supportedlock		
current-user-privilege-set		
acl		
acl-semantic		
owner		
displayname		PedigreeTutorial.doc
getcontenttype		application/msword
getcontentlength		60416
getcontentlanguage		en
getetag		908497668_48563_-1
creationdate		2002-06-04T14:24:26Z
getlastmodified		Tue, 04 Jun 2002 14:24:26 GMT
date.issued	dc	2002-06-11
date.available	dc	2002-06-11
type	dc	text
date.created	dc	2002-06-11
subject	dc	pedigree; tutorial; Dublin Core
creator	dc	Carmen Pancerella
date.modified	dc	2002-06-11
date.valid	dc	2002-06-11
publisher	dc	CMCS
description	dc	This is the CMCS pedigree tutorial.
title	dc	CMCS Pedigree Tutorial
contributor	dc	Jim Myers; Brett Didier; Eric Stephan; Tom Allison

dc = Dublin Core

Add Delete Save Close



GRIMech Data on CMCS Slide Server Has Pedigree Metadata



DAV Explorer

File Edit View Help

View/Modify Properties

cmcs.ca.sandia.gov:18081/slide/files/projects/grimech/thermo/c2h.xml

Tag	Namespace	Value
supportedlock		
supported-privilege-set		
acl		
owner		
contributor	dc:	Michael Frenklach
relation.references.href	dc:	http://cmcs.ca.sandia.gov:18081/slide/files/projects/grimech/thermo/c2h.xml
description	dc:	GRI-Mech thermodynamic polym...
subject	dc:	
relation.references.relativeHref	dc:	../description/nasa_thermo.xml
displayname		c2h.xml
getcontenttype		text/xml
getcontentlength		769
getcontentlanguage		en
creationdate		2002-06-03T21:54:08Z
date	dc:	May 2002
creator	dc:	GRI-Mech team
getlastmodified		Mon, 03 Jun 2002 22:10:14 GMT
source	dc:	GRI-Mech 3.0
title	dc:	C2H thermodynamic polynomials

Add Delete Save Close

GRIMech thermo data is annotated with pedigree information, including bibliographical references.



Pedigree Browser Portlet



Collaboratory for Multi-Scale Chemical Science

About CMCS
Portal
Project Team
Publications
Contacts

Home CMCS Search Pedigree Data Insertion

CMCS Pedigree Browser Portlet

Resource

Creator

Reference

Contributors

Resource Name

Subject and Keywords

Description or Abstract of Resource

EMSL Computational Results DataBase V1.9 11/11/2000 - 255 Molecules, 41 Atoms (37,071 entries)

Select from this column first!

☒ Molecular Structure
(internal coords)

☐ Cartesian Coordinates

☐ Harmonic Vib. Freqs

ScaleFactor:

☐ Scale HF only

☐ Total Energies

Geometry:

☐ Relative Energetics

Larger Molecules ☐ All

H2 1-Sigma g+

LiH 1-Sigma +

BeH 2-Sigma +

BeH2 1-Sigma g+

BeO 1-Sigma +

BeOH 2-A'

BeF 2-Sigma +

BeS 1-Sigma +

BeCl 2-Sigma +

BH 1-Sigma +

BH+ 2-Sigma +

BH2 2-A1

BH3 1-A1' borane

B2 3-Sigma g-

Errors

With Respect To

Basis Set

The

A reference URL is stored as pedigree metadata for the EMSL Computational Results Data on the CMCS Slide server. Pedigree browser allows this URL to be retrieved with a click.